# Sport policy evaluation: what do we know and how might we move forward?

Chen, Shushu

*DOI:*
10.1080/19406940.2018.1488759



*Document Version*
Peer reviewed version

*Citation for published version (Harvard):*
Chen, S 2018, 'Sport policy evaluation: what do we know and how might we move forward?', *International Journal of Sport Policy and Politics*. https://doi.org/10.1080/19406940.2018.1488759

*Paper accepted for publication in the International Journal of Sport Policy and Politics

# Sport policy evaluation: What do we know and how might we move forward?

**Shushu Chen**

**School of Sport, Exercise and Rehabilitation Sciences, University of Birmingham, s.chen.5@bham.ac.uk**

## Abstract

Evaluation, a product of the movement for evidence-based policy, is a key step in policy cycles. However, many studies existing in sport policy literature have failed to address underpinning methodologies in a rigorous manner and to provide justification for the use of certain measures handpicked by evaluators. As yet, no study has explicitly reflected on the value of evaluation or has systematically discussed how mainstream evaluation theories have been used in sport policy studies. Such articulation is necessary in order to provide researchers with additional resources for making informed and strategic methodological choices and to ensure the quality of their analysis. Thus, this paper discusses the development of evaluation in general, and examines existing literature on sport policy evaluation. It then goes on to outline four especially noteworthy public policy evaluation frameworks: experimental design, constructivist evaluation, utilisation-focused evaluation, and realist evaluation. Next, it uses a specific example to highlight the strengths of realist evaluation as a tool for unpacking additionality and understanding the logic of theory. In conclusion, the paper suggests using theory-based evaluation

frameworks (specifically, realist evaluation) to inform sport policy failure or success for future sport evaluation research.

**Keywords**: Policy evaluation, sport, realist evaluation, theory-based evaluation

## Introduction

Evaluation is one of the key components in policy processes or cycles (Easton, 1953; Hill, 2005; Hogwood & Gunn, 1984). It plays an important role in decision-making processes (Foley, 1992; O'Brien, 2013) by providing policymakers with accounts of initiatives' impacts (Weiss, 1993). The last couple of decades have seen the rapid development of policy evaluation research in many fields (e.g., Furubo, Rist, & Sandahl, 2002; Jacob, Speer, & Furubo, 2015; Mastenbroek, van Voorst, & Meuwese, 2016), such as nursing, housing, education, medicine, engineering, social services, and international development.

Especially in the UK, considerable impetus was given to policy evaluation in the late 1990s following the evidence-based policymaking movement: New Labour was elected in 1997 with a manifesto that said 'What matters is what works' (Cabinet Office, 1999). A few years later, against this background of an evidence-based policymaking movement, a call for evidence-based sport policy evaluation was made explicit in the national sport policy document Game Plan (DCMS, 2002), which suggested strengthening sport policy evaluation to 'enable policy-makers to construct and target effective interventions' (p. 79). Following on from that, a number of sport policy studies have been commissioned by government departments and agencies concerned with promoting accountability and control (HallAitken, 2009; Loughborough Partnership, 2009; Grant Thornton et al., 2011a, b, 2012, 2013).

Alongside this growth in sport policy/programme evaluation, there has been increasing concern over evaluation quality (Coalter, 2017; Henry, 2016; Weed, 2014). De Bosscher et al. (2011) also noted how little consensus has emerged in the field of sport as to what constitutes policy effectiveness and how that is best measured. To overcome such challenges, Houlihan (2011) suggested that studies of sport policy evaluation could draw from a broad range of theoretical and methodological approaches used in other public policy areas and argued that 'there are lessons to be learnt from the extensive evaluation work being conducted in similarly complex policy areas' (p. 558), such as policing and health.

In evaluation research, overall, Stufflebeam and Coryn's (2014) review of existing work published between 1930 and 2014 summarised that there are approximately twenty-three unique evaluation approaches available. So far, however, there has been little discussion about how one might choose wisely from available options and why one evaluation framework is more appropriate than another for sport policy/programme evaluation.

In acknowledging both the criticisms faced in the field of sport policy/programme evaluation and the need to build a useful knowledge base of mainstream evaluation frameworks, this paper seeks to provide a critical review of the theoretical bases for some of the most commonly used approaches found in mainstream literature – experimental evaluation, constructivist evaluation, utilisation-focused evaluation, and realist evaluation – and to discuss their strengths and weaknesses. Particularly, we put forward our argument that use of the realist evaluation framework can contribute to the informing of policy and practice. The paper then examines one specific example of such an approach

being applied. A conclusion, which summarises some difficulties and future agenda associated with sport policy evaluation, is provided at the end.

It is important to note at the outset that the primary focus of our review and the further discussion pivot around *impact evaluation* (i.e., appraisal regarding the acceptability and effectiveness of practices designed to improve sport and physical activity along with their associated social impacts,  House of Sport, 2013), rather than e*conomic impact evaluation* (which is typically assessed using economics language and endorses the use of cost-benefit analysis, HM Treasury, 2003).

## The development of evaluation

Policy evaluation, as defined by Gerston (1997, p. 120), is to assess 'the effectiveness of a public policy in terms of its perceived intentions and results'. Vedung (1997, p. 3) expanded this definition and described policy evaluation as the 'careful retrospective assessment of the merit, worth, and value of administration, output, and outcome of government interventions, which is intended to play a role in future practical action situations'. Put simply, evaluation provides a retrospective analysis of a project, programme, or policy to assess how successful or unsuccessful it has been and what lessons can be learnt for the future (HM Treasury, 2003).

Much of the research work related to evaluation originated in the United States (Nutley & Webb, 2000) and derived from the area of education (Stufflebeam, Madaus, & Kellaghan, 2000). Tyler (1942), Cronbach (1963), Guba (1969), Scriven (1967), Glaser (1963), Stufflebeam (1966), and Campbell and Stanley (1966) are pioneers in the field. However, Nutley and Webb (2000) pointed out that American evaluation research often focuses more than European evaluation research does on single-issue and short-term

effects, rather than on policy planning, because of the USA's fragmented and decentralised political system.

A number of journals and publications are dedicated exclusively to evaluation research, and among these are *Evaluation and Program Planning*, *Studies in Evaluation*, and *Evaluation Review*, and the *American Journal of Evaluations* appeared first, in the 1970s; these were followed by *Research Evaluation*, *Evaluation: The International Journal of Theory*, *Research and Practice* and *the Evaluation Journal of Australia* in the 1990s.

The key position of evaluation in the policy cycle process was identified long ago (Easton, 1953; Hill, 2005; Hogwood & Gunn, 1984). The need for policy evaluation seems to stem from the pressure to illustrate some kind of return on a state's investment (Hood, Dixon, & Wilson, 2009) as well as from the limits of our understanding regarding how government interventions work and the subsequent effects (Hogwood & Gunn, 1984); all of this makes it necessary to monitor and evaluate policy (Hogwood & Gunn, 1984).

The role of policy evaluation was once seen as being to contribute towards predominantly *ex post facto* assessment of the extent to which policy interventions achieved their intended effects (Sanderson, 2000). That is, policy *outcomes* were more often to be evaluated than the implementation process was; the evaluation of the implementation process was referred to as *process* evaluation (Daugbjerg et al., 2009). This goal-oriented evaluation view seems logical if policy is goal driven and can complete the policy cycle by providing feedback to improve policy (Colebatch, 1998; Parsons, 1995). Weiss (1993) nevertheless warned us that evaluation is 'a rational

exercise that takes place in a political context' (p. 94). As most policy evaluations are commissioned by government departments and agencies (Sanderson, 2000), a top-down-oriented evaluation suffers from criticisms such as lack of accountability and overly controlled research findings (Henkel, 1991; Sanderson, 2000). Other scholars have also raised concerns over conducting policy evaluation based solely on goal attainment, which fails to provide insights into how effective policy interventions have been or into which specific factors have contributed to reaching policy goals (De Bosscher et al., 2011; Suomi, 2004). There has therefore been a call for the use of process evaluation to complement outcome evaluations, as it has particular value for multisite measurement whereby the same interventions are delivered and received in different contexts (Mansfield, Anokye, Fox-Rushby, & Kay, 2015).

Subsequently, policy evaluation focusing on process analysis has emerged (most notably the theory-driven process evaluation approach, Chen, 1990, 2015), aimed at informing policy and resource allocation decisions (Foley, 1992; O'Brien, 2013; Tilley, 2000). The major strengths of the process evaluation approach are that this approach helps to identify information critical to understanding how a programme is implemented; as such, this careful understanding of the implementation process contributes to the dissemination of the programme to other settings (Chen et al., 2008).

The fulfilment of both functions (outcome and process evaluations) has not been very successful, however, hampered by the complex nature of policy systems (Sanderson, 2000) and deficient methodology (Rist, 1995; Weiss, 1993). Summarising his edited collection on policy evaluation, Rist (1995) noted a range of methodological issues faced in policy evaluation research, including misunderstanding and misapplication of the

measurement indicators, for example, and disagreement on the boundaries within which one would look for evidence of utilisation.

In addition to concerns over research quality, there is also the issue of political influence (Weiss, 1993; Weiss & Bucuvalas, 1980), whereby the so-called evaluation evidence is perceived to be useful when it supports and reinforces policy beliefs or current policy commitment (Coalter, 2017; Gray & Jenkins, 1995; Rowe, 2005; Weiss, 1993). Politically-constrained evaluations are thus limited in their ability to contribute to policy learning in a real sense (Lindsey & Bacon, 2016). On the one hand, in-house or commissioned evaluations may be subject to accusations of 'confirmation bias' (Nickerson, 1998), as evaluators are under political pressure to look good and to act upon related findings (Schoenefeld & Jordan, 2017; The Lse Gv, 2014; Weiss, 1993). Osterlind (2016) also questioned the state-appointed commission type of inquiry in sport policy, as this approach often tells us more about only certain political values and objectives than about real challenges. On the other hand, the usefulness of independent evaluations has been contested by several authors (Chelimsky, 2006; Erik, 1995; Patton, 1997; Weiss, 1993) claiming that these evaluations often fail, ultimately, to facilitate evaluation knowledge in the policy process and to properly inform policy. As noted by Rist (1995), whether and when policymakers might use evaluation data depends upon 'the circumstances of the kinds of questions being asked, the time frame in which the answers have to be found, the trust the policy maker has in the sources of the data, the ready availability of the evaluation data, and on and on' (p. xvii).

An important concept in policy evaluation is worth elaborating upon here, namely *additionality* (Foley, 1992; Storey, 1990). As explained by Rist (1995), teasing out

additionality – identifying which changes in conditions have been caused by the efforts of a policy or programme and which ones have not – remains one of the most difficult methodological challenges facing a policy evaluator. Without assessing additionality, it is not clear what the intervention/policy is contributing over and above what would have happened anyway. The results of the evaluation might therefore provide a misleading picture of the value of the international/policy. The term additionality and its associated identification of the counterfactual scenarios have been discussed in detail elsewhere (Chen, Henry, & Ko, 2013), but it is worth briefly revisiting the development and application of the respective concepts here.

The concept of additionality originally came from the evaluation of innovation and technology policy for the purpose of assessing public funds made available in addition to those that would have been provided by default (Buisseret, Cameron, & Georghiou, 1995). In the UK, particularly, additionality has become a key concept in appraising proposals and evaluating policy initiatives, together with other concepts such as general impacts, effectiveness, efficiency, and value for money (HM Treasury, 2003). The first step in calculating additionality is to set out the counterfactual scenario, which means establishing what would be the case if the antecedent in question were not true or, in other words, what would have happened if the intervention had not gone ahead.

English Partnership (2008) offered practical guidance on how to take into account the additionality of intervention for the purpose of ensuring net impact assessment. In brief, estimation of the net impact involves consideration of four key concepts: *leakage* (the extent to which the gross impact of benefits generated and intended for a particular group, region, or country incorporates beneficiaries from other groups, regions, and

countries), *displacement* (where a new provision displaces other activities or services previously provided), *substitution* (when consumers of a service or beneficiaries of an intervention simply substitute a new service provision for one they previously used or benefited from), and *multiplier effects* (the extent to which direct benefits from an intervention trigger further additional indirect benefits). Nevertheless, the difficulties encountered when assessing additionality in evaluation studies have been acknowledged by McEldowney (1997), and we will turn to this point later.

Evaluation theories that have been applied in the mainstream literature include, for example, formative and summative evaluation (Scriven, 1967); case study evaluation (Yin, 1992; Keen & Packwood, 1995;); context, input, process, and product evaluation (Stufflebeam, 1967); participatory evaluation (Cousins & Earl, 1992); and meta-evaluation (Scriven, 1969; Stufflebeam, 1978). In a later section, we focus on introducing the following four approaches that are commonly used in evaluation research (Stufflebeam & Coryn, 2014) and that represent four main perspectives on evaluation (Pawson & Tilley, 1997): (quasi-) experimental design, constructivist evaluation, utilisation-focused evaluation, and realist evaluation. It is important to note here that we sought to compare the merits of the four approaches but not necessarily to review how useful or effective they were perceived to be by policy makers[1].

## Evaluation in the field of sport: methodological consideration

The role that sport policy evaluation is considered to play, at least in the EU, is similar to that of other mainstream public policy evaluations – to make for evidence-based policymaking (European Commission Communication, 2011). As suggested by the European Commission (2011, p. 8), ‘policy-making to implement the sport provisions in

the Lisbon Treaty needs a sound evidence base, including comparable EU-wide data on social and economic aspects of sport'.

In recent years, although there has been an increasing amount of literature both directly and indirectly concerned with sport policy/programme evaluation (e.g., Chen & Henry, 2016; Coalter, 2013; Girginov, 2016; Weed, 2014), policy evaluation in the context of sport is still in its infancy. It was indicated in Daugbjerg et al.'s (2009) content analysis of 27 national policy documents relating to physical activity promotion in the EU that, although most of the respective policies emphasised the importance of an evaluation plan, only about half of the policies indicated an intention for evaluation focusing on outcomes rather than on the implementation process. This is in contrast to other sectors, e.g., education and environmental and climate policies, which are considered to be well-established disciplines for evaluation and which undergo evaluation both of outcomes and of processes (EEA, 2016; Stufflebeam & Coryn, 2014).

Henry (2016) shared his views on the issue of inadequate evaluation research in the field of sport. Drawing conclusions from his involvement with the meta-evaluation of London 2012 Olympic and Paralympic sport legacy programmes, Henry (2016) indicated that some sport programmes had not yet been subject to any evaluations. Within those evaluations that had been carried out, reporting tended to be limited to attendance or participation figures (e.g., the Gold Challenge programme).

There is a consensus in the field that the evaluation research design is still insufficiently rigorous (Coalter, 2017; Henry, 2016; Houlihan, 2011; Jones, Edwards, Bocarro, Bunds, & Smith, 2017; Weed, 2014). Of some noteworthiness, specifically, is the failure of many sport policy evaluation studies and programme-evaluation studies,

which are often commissioned ones, to adopt any explicit theoretical framework (e.g., the Sport Maker evaluation); a range of approaches (e.g., self-completion of questionnaires and case studies) are handpicked and grouped together by evaluators (often under the influence of stakeholders) without there being any justification for the respective methods. As Houlihan (2011, p. 557) concluded,

> Studies of the impact of sports development initiatives abound with examples of the selection of inappropriate indicators, inconsistent measurement and dubious generalisation. The failure to take on the challenge of effective policy evaluation allows the critics of the policy to argue that the absence of evidence equates to the absence of impact.

Similarly, Lindsey and Bacon (2016) reviewed research reports undertaken to monitor and evaluate youth sport and physical activity initiatives in England between 2002 and 2010; their review concluded that those research reports 'tended to describe the implementation of data collection methods without demonstrating any consideration for broader research methodologies' (p. 84).

In some of his work, Coalter (2007, 2010, 2011, 2013) highlighted the difficulty of conducting evaluation research in the field of sport, where conceptual and measurement approaches tend to agree relatively less than they might in medical or physical research, for example. He specifically noted several methodological failures from his years of experience compiling the Value of Sport Monitor. These failures included, for example, problems with conceptual variety, methodological weaknesses (e.g., self-reporting), and failure to address issues of conditions. Particularly, self-completion survey approaches seem to be used as a common programme evaluation practice in the field of sport (Henry, 2016). For example, the Department for Education conducted annual surveys (2003–2010) of sport participation, in which the number of

children taking part in sport and physical education is recorded by in-school staff (Henry, 2016). Looking at the School Sport Partnerships programme, Smith and Leech (2010) found that self-completion of the Physical Education, School Sport, and Club Links surveys, led by the Department for Education and Skills and the Department for Culture, Media, and Sport, were used by programme representatives as part of the exercise for monitoring and evaluating the effectiveness of their work.

Regarding methodological rigour, one relevant issue is that sport policy evaluation – and, more specifically, programme evaluation – often fails to thoroughly assess additionality. For example, as noted by Henry (2016), most of the London-2012-related project evaluation did not assess the policy counterfactual scenario (which policies and initiatives would have been used if the Olympics had not been hosted in London). Using the evaluation of the Free Swimming scheme (PricewaterhouseCoopers, 2010) as an example, Henry (2016) commented that this evaluation constituted the *only* project evaluation which incorporated a full account of additionality elements.

Such a disregard for additionality seems to have been a common issue facing sport policy/programme evaluation. Weed (2014) concluded from his research analysing London 2012 tourism legacy strategy that positive tourism legacy outcomes had been claimed without consideration as to whether they were additional or attributable. He also highlighted the importance of assessing additionality and attribution, and he applied this in his London-2012-related discussions (Weed, 2010, 2014), suggesting that directly attributable effects should be measured in addition to opportunity costs. Such perspectives reflect a growing realisation that there is a need for theory-based evaluation (Armour, Sandford, & Duncombe, 2013; Bailey et al., 2009; Coalter, 2017; Henry, 2016;

Houlihan, 2011; Jones et al., 2017; Wells & Arthur-Banning, 2008) to 'de-mythologise' which sports work, for which subjects, in which conditions, and why this is so (Coalter, 2017, p. 146). Coalter (2017) argues that

> There is now a widespread acknowledgement of the need for an understanding of programme processes – the nature of participants' experience and the mechanisms which explain any measured changes in values, attitudes or behavior. We have limited understanding about what sports and sports' processes produce what outcomes, for which participants and in what circumstances. (p. 147)

In parallel, from sport policymakers and governing bodies' side, the generally disappointing quality of sport evaluation has received attention (Sport England, 2016). For the purpose of fulfilling a commitment made in the latest national strategy plan to go beyond simply collecting numbers, Sport England – a non-departmental public body that sits under the DCMS in the UK and is responsible for the development of mass sport participation – has recently launched a standard evaluation framework to assist the measurement of funding streams and projects with a view to enhancing evaluation quality (for more detail, see: https://evaluationframework.sportengland.org/). Sport England advocates a more rigorously defined evaluation being adopted by all Sport-England-funded programmes, recommends the employment of not only outcome/impact evaluation but also process evaluation, and suggests that programme theories be outlined and presented explicitly by developing logic models. In Northern Ireland, a similar commitment has been made by Sport North Ireland to evaluate the impact of sport policy, through a number of programme evaluations, to justify "investment and improving the evidence base for 'what works' and most importantly 'why'" (House of Sport, 2013, p. 5).

Finally, the amount of funding for evaluation research activity varies considerably across public service areas. One may argue that the lack of attention to evaluation in the

field of sport and the relatively poor evaluation quality mentioned above are a direct result of the financial constraints of the field (in contrast with other areas, such as health care, education and criminal justice, that consume high levels of government recourse and political attention, Nutley, Walter, & Davies, 2007). Evaluation quality might also be compromised when there are so few specific policy 'customers' to be found (Davies, Nutley, & Smith, 2000, p.357), and researchers in many cases await invitations to tender for particular projects. Such an evaluation research environment is unhealthy, in some extreme cases, sport policy evaluation might at times become user-led (i.e. government-led) research that merely reinforces existing political ideologies rather than questioning policy intentions. However, this is not to neglect that, in real-word evaluations, budget, time, and data constraints are also common issues in other fields' evaluation research (Bamberger, Rugh, & Mabry, 2012). To overcome these constrains, Bamberger et al. (2012) offered a seven-step approach to ensure maximum possible methodological rigour within a particular evaluation context. To reduce costs, they suggested that evaluators can make greater use of secondary data, revising the sample design and size and streamlining data collection and analysis, without necessarily compromising the quality of the evaluations.

## Four approaches for evaluation

### *Experimental and quasi-experimental design*

Campbell and Stanley's (1963) classic OXO notation is well-known in the evaluation field. The OXO notation, referring to experimental designs, describes an experiment where participants are randomly assigned to two equivalent groups for comparison: one treatment group (applied with an intervention) and one control group. The impacts of the

intervention can be estimated by comparing the outcome measures from these two groups. A standard experimental evaluation process starts from the design of methodological strategy. It normally begins with establishing a philosophical grounding for answering the question ‘what can be known’, and the ways in which knowledge of social reality can be obtained and judged as being both adequate and legitimate epistemology - this is the theory of ‘causal explanation’.

Cook and Campbell (1979) argued that the randomised control trial designs have the ability to eliminate threats to internal validity and are often considered as the ‘gold standard’ for assessing impact in biomedical research and medicine (Cartwright, 2007). Similarly, the notion of an ‘evidence of hierarchy’, with randomised controlled trials at the top (Hadorn et la., 1996; Leigh, 2009), seems to suggest that such an approach is superior to other methods. As a result, such approaches have received prominent attention in countries like the United States and the United Kingdom. For instance, the U.S. government specifically required use of randomised controlled experimental design to evaluate federally funded innovations in education and other social services during the late 1960s and early 1970s (Stufflebeam & Coryn, 2014); similarly, the experimental approach gained hegemonic status in fields such as healthcare and education in the UK (Davies & Nutley, 2000; Davies, 1999). Proponents of the experimental evaluation approach consider that such an approach provides rigorous and credible evidence about whether an intervention is effective (Cook, 1995; Shadish, Cook, & Campbell, 2002). Later on, after the introduction of the experimental design, quasi-experimental designs emerged to overcome the difficulty of applying randomised control trials in real word

evaluation environments (Rossi, Lipsey, & Freeman, 2004). This newer approach has similar features to experimental designs but does not use random assignment.

Amongst those sport-related studies whose theoretical frameworks of evaluation have been clearly specified, we found that sport psychologists, sport scientists, and pedagogues tend to be in favour of using an experimental approach to examine, for example, the impact of a particular sport programme on young people's psychological and physical development (Gabriel, DeBate, High, & Racine, 2011), the impact of a physical education intervention on promoting participation in physical activity and movement skill proficiency among adolescents (Dudley, Okely, Pearson, & Peat, 2010; Gortmaker et al., 1999), and the impact of a community sport intervention on physical activity participation levels (Mansfield et al., 2015).

Despite their prominent positions in the field of evaluation, the experimental and quasi-experimental method-based approaches to evaluation have been criticised especially by scholars coming from the qualitative camp (Chen, Donaldson, & Mark, 2011) who consider that the process of experimental evaluation is unable to provide any real insight into the underlying causal mechanisms that produce treatment effects. This issue is often referred to as the 'black box' problem (Scriven, 1991): The utility of an approach can reveal whether an intervention has an impact on outcomes, but the results of experimental studies do not reveal under which conditions and through which configuration of factors certain outcomes are achievable.

The second major criticism of experimental evaluation is that the approach fails to account for the power of contextual influences over an intervention (Cronbach, 1982; Greene, 2009; Pawson & Tilley, 1997). For example, in Gabriel et al.'s (2011) evaluation

of a youth sport programme designed to promote beneficial changes to physical and psychological assets in girls from 35 elementary schools, although the study noted the strength of using quasi-experimental study design in identifying the programme's effects, the influence of context (e.g., school environment and facilities) was not appreciated. In addition, as Gabriel et al. (2011) recognised, different programme facilitators (school teachers or research staff) may have differentially affected participants' perceived coercion, which thus led to increased social desirability bias. As a result, key findings emerged from the evaluation are often summarised and lifted out of context (Sanderson, 2000).

Another major criticism of experimental evaluation concerns the low level of feasibility and validity associated with employing experimental design in the context of a social system (Coalter, 2017; Sanderson, 2002; Weiss, 1997). In practical cases, the social world is *complex*, *open*, and *dynamic* (Pawson & Tilley, 1997). Conducting an experimental evaluation in such contexts is challenging, as it is hard to make sure that two experimental groups are identical at the outset, and it is difficult to insert an intervention clearly into the 'experimental group' without influencing the 'control group' (Scriven, 2008). In Dudley et al.'s (2010) study, due to the fact that the implementation of a school-based physical activity programme was delivered in a single school, a potential issue was that participants from the intervention group and the control group may have talked to each other, potentially affecting internal validity. As noted by Stufflebeam and Coryn (2014), providing a neat experimental comparison seems virtually impossible in most practical contexts, particularly in the real world of schools; they further explained that 'it is often difficult to arrange and sustain treatment and control

groups and convince administrators and parents and other stakeholders that the benefits of such procedures are worth the difficulties of keeping groups separate and sustaining their different treatment experiences' (p. 252). Since we are rarely dealing with a lab-based experiment, it is not easy to draw the conclusion that a specific intervention contributes to certain outcomes.

***Constructivist evaluation***

In a rejection of the principles and procedures of the randomised controlled experimental design, Guba and Lincoln (1989) developed the tenets of constructivist evaluation. Having recognised the complex processes of human understanding and interaction, constructivists Guba and Lincoln (1989) believe that, because knowledge of the social world is socially constructed, quantitative approaches do not seem to offer much potential for improving policymaking or for meaningful involvement by stakeholders in evaluation. This view drove the development of evaluation forward from a focus on *outputs* to one on *processes*. The primary distinguishing characteristic of the constructivist evaluation approach is that it appreciates the existence of different stakeholders, it emphasises the need for the evaluator to explore how different stakeholders interpret a problem or an intervention, and it tries to bring stakeholder and evaluator together to understand each other.

The key principles provided by Lincoln and Guba (1985) for guiding evaluators in this type of research are, firstly, that it is important for evaluators to have open minds and to be willing to learn the circumstances surrounding different perceptions of reality. Secondly, it is important for evaluators to experience the context within a programme and to discover how different stakeholders interact with each other. Consequently, the

research design and the strategy of data collection used for constructivist evaluation allow the evaluation approach to emerge or unfold as the research progresses, whereas the traditional experimental evaluation paradigm abides by the rules that existing theory should guide the data collection process.

Constructivist evaluation starts with identifying key stakeholders, and it focuses on understanding humans' meanings. As all stakeholders are involved in this process, they start to establish their 'constructions' about a programme, they negotiate between themselves, and they assemble their claims, concerns, and issues in order to produce consensual constructions. This unique progress of constructivist evaluation determines that the utility of a constructivist evaluation is regarded as an attempt to reconsider unresolved constructions through a repetitive process of thinking and discussion until enlightenment/consensus is obtained.

In the field of sport, as Kay (2009) noted, the potential for evaluating sport-related issues by adopting a constructivist perspective has not yet been properly explored. While Girginov and Hills' (2009) study of Olympic sports development legacy viewed this approach as being useful for enabling 'locally meaningful experiences to feed into politically and practically useful policies' (p. 178), there are some disadvantages associated mainly with the openness and the exploratory nature of constructivist perspectives (Stufflebeam & Coryn, 2014). The need for full stakeholder involvement and ongoing interaction inevitably requires an extensive and time-consuming evaluation process (Stufflebeam & Coryn, 2014). Moreover, as noted by Pawson and Tilley (1997), constructivist evaluation fails to recognise the asymmetries of powers that may exist

between different stakeholders, so that it is difficult to reach a joint construction of claims by various stakeholders who have totally opposing views about an intervention.

***Utilisation-focused evaluation***

Beginning around the late 1970s, evaluators faced an identity crisis, and there were debates concerning how evaluation could be made more useful, feasible, and effective (Stufflebeam & Coryn, 2014). A different theoretical perspective – utilisation-focused evaluation – emerged, led by Patton (1984, 1997, 2008), who believed that an evaluation had to be judged by its usefulness. Theoretically, this type of evaluation is aligned with relativistic and constructivist perspectives in terms of the selection and application of values – that is, utilisation-focused evaluators should seek consensus on both values and judgement to support the decision-making process (Stufflebeam & Coryn, 2014). Following Peirce's (1931) theory of pragmatism, Patton suggested that evaluation research did not have to follow certain epistemological axioms; rather, he considered that evaluation research was best learned through exemplars, and it should be concerned with whether the practical cause of policymaking is advanced. As evaluation evolves to become more useful and feasible, evaluators increasingly take a more feet-on-the-ground view of methodology.

Utilisation-focused evaluation is often recognised as being a comparatively useful form of evaluation, one which helps to narrow the gap between generation of evaluation information and application of evaluation information to programme decision making and improvement (Stufflebeam & Coryn, 2014). Supporters of pragmatic evaluation (for example, Alkin, 2004; Cronbach & Associates., 1980; Stufflebeam, 1966; Weiss, 1972) have argued that an evaluation's impact should be substantial, meaningful, and relevant.

In this respect, reviewing the historical emergence of different programme evaluations provides an overall framework within which individual evaluators can proceed to develop evaluation designs unique to the circumstances they encounter. This exercise helps with targeting evaluation more carefully and also increases the likelihood of evaluation findings being put to good use by policymakers.

The utilisation-focused method places emphasis on the final results rather than on rules, and this represents a difficulty in terms of generalisation and of deciding why and how to do it. As noted by Alkin (2004), a utilisation-focused evaluation 'is concerned with designing evaluations that are intended to inform decision-making, but it is not their only function to ensure that evaluation results have a direct impact on programme decision-making and organisational change' (p. 45). The crucial point here is that evaluators must determine their studies according to the intended evaluation uses and must focus the studies in the correct way so as to produce findings that an identified group of intended users can and probably will value and will apply to programme improvement (Stufflebeam & Coryn, 2014).

Despite the fact that pragmatic evaluation makes a great effort to emphasise the advantage and necessity of focusing on utilisation, its close association with evaluation users is perceived to be its main limitation. Such an approach seems to be vulnerable to bias and corruption by user groups (Ramirez & Brodhead, 2013), namely the programme stakeholders, who are likely to exercise much control over what is examined, what questions are addressed, and what information is used to address the questions. To conduct evaluations, evaluators can only follow instructions which are provided by this

user group – this leads to the evaluation results suffering from bias[2] (Pawson & Tilley, 1997; Stufflebeam & Coryn, 2014).

***Realist evaluation***

The emergence of theory-driven evaluation (Chen & Rossi, 1980, 1983, 1987) during the 1980s sought to address several limitations (e.g., inconsistent results, nonreplicability) associated with the method-driven evaluation theories reviewed above. As illustrated by Chen (1990), this approach places emphasis not only on the implementation and effectiveness of the intervention, but also on the causal mechanisms and the contextual factors underlining the changes. Realist evaluation[3] belongs to the family of theory-based perspectives on evaluation which stress that an evaluation has the task of testing out the underlying programme theories. It aims to unpack 'what works, for whom and in what circumstances' (Pawson & Tilley, 2004, p. 1). In other words, outcomes are explained by the actions of specific mechanisms in specific contexts, and this explanatory structure is established by a combination of theory and experimental observation. Thus, central to the whole realist evaluation framework is the concept of 'theory of change', i.e., there is a causal relationship between 'outcomes' and 'mechanisms' acting in 'contexts' in any evaluation project. As highlighted by Pawson and Tilley (1997, p. 57), 'programmes work (have successful outcomes) only in so far as they introduce the appropriate ideas and opportunities (mechanisms) to groups in the appropriate social and cultural conditions (contexts)'. This formula is termed the Context–Mechanism–Outcome (CMO) configuration.

The arrival of realist evaluation theory stems from the trend for lack of certainty in evaluation research, where carefully designed policy intervention works only

occasionally and the reported results are often inconsistent (Pawson & Tilley, 1994). Pawson and Tilley (1997) claimed that 'methodological failure is at the root of the capriciousness of evaluation research' (p. 292). They (1997) argued that the (quasi-experimental) evaluation design constituted a 'fine strategy for evaluating the relative performances of washing powders or crop fertilizers' but 'a lousy means of expressing the nature of causality and change going on within social programmes' (1997, p. 292).

Following on from Merton's (1968) middle-range theory, Pawson and Tilley (1997) sought to explain how an intervention leads to a particular outcome under the strategy of 'scientific realist' analysis[4] (Pawson, 2013, p. xviii). They emphasised that 'evaluation should follow a *realist* methodology' (Pawson & Tilley, 1997, p. xiii); specifically, they attempted to conduct evaluation with a depth realist ontology using the logic of realist explanation for uncovering certain 'causal powers' of individuals and institutions' (Pawson, 2000, p. 294). In contrast to the evaluation approaches mentioned above, realist evaluation recognises that the evaluation of any social evaluation programme is undertaken within a particular social system – in this case an *open* system, a type of system which is active and alterable by those with reasoning and resources. Thus, Pawson and Tilley (1997) have argued that the complexity of human societies and the open system within which interventions are conducted further demand the examination of causation (Pawson, 2013); and only after seeing a complete picture of the complex, differentiated, and intertwined programmes can a more appropriate and realist implication for research and evaluation be provided.

The preceding sections of the review have identified some existing methodological issues associated with sport policy/programme evaluation, subsequently

highlighting that there is a need for sport evaluation research to be grounded in sound theory; to move away from merely outcome-based evaluation towards a combination of process-based evaluation and outcome-based evaluation, in order to identify which factors of a sport policy intervention have caused particular outcomes in given contexts.

We argue that key principles underlying Pawson and Tilley's (1997) realist evaluation framework seem to offer some solutions for addressing the aforementioned issues and for contributing to the development of methodological rigorousness. Beyond simply identifying inputs, throughputs, outputs, and outcomes, such an approach has major methodological implications: It seeks to develop explanatory structure whereby specific outcomes can be explained by the action of specific mechanisms in specific contexts. This is particularly helpful for unpacking the implicit and explicit assumptions underpinning a particular sport policy or programme, assumptions that may or may not be widely shared, understood, or agreed.

Although realist evaluation has modest popularity in mainstream literature areas (Marchal, van Belle, van Olmen, Hoerée, & Kegels, 2012), its application in the field of sport is still limited and young (Bell, 2004; Chen & Henry, 2016, 2017; Girginov, 2016; Harris, 2018; Hughes, 2013). We cite here a recent study (Chen & Henry, 2016) that has applied realist evaluation to illustrate the merits of the approach as well as to demonstrate how the key concepts have been employed.

Against the background of London hosting the 2012 Olympic and Paralympic Games, the evaluation examined the impact of a programme, namely *Workplace Challenge*, on leveraging sport and physical activity participation and explored the possible impact that the 2012 Olympic Games may have had on the promotion and

operation of the programme. Workplace Challenge was a sport and physical activity programme that aimed to stimulate competitions between local organisations in terms of the recorded levels of sport and physical activity undertaken by employees over a given period.

The programme theory of Workplace Challenge was uncovered from document analysis and interviews with stakeholders, which contributed to the development of CMOs, coupled with quantitative data gathered from the programme participants. Going beyond simply reporting the number of organisations and employees taking part in the programme and testing the statistical significance difference in sport and physical activity participation before the programme versus afterwards, efforts were made to identify which mechanisms worked for whom and under which circumstances they worked. Additionally, to further assess the event's impact on stimulating participants' interest in programme participation, a series of London-2012-related questions were included in the surveys and interviews, and the corresponding data were subsequently used in the construction of CMOs.

This allowed different theories of programme impact (London 2012 related and programme related), linked with their respective conditions and causal mechanisms, to be presented clearly. Two CMOs were developed based on two hypotheses (see Table 1): (a) participants with different exercise intensity levels[5] at the start of the programme might react differently to mechanisms, thus generating different outcomes; (b) programme participants from different types of work organisations might engage with the programme in different ways and to different degrees, as facilitated or restrained by the nature and contextual features of their organisations.

[Insert Table 1 about here]

By developing the two CMOs, the process has helped to formulate our understanding of *how* the Workplace Challenge programme was structured and designed to *work* for *certain* specific groups of participants and, in the cases of groups or organisations for which it worked less effectively, of subsequent changes that could be made in the delivery of the programme. To summarise some of the key messages:

- *Which of the generative mechanisms in place worked to foster positive outcomes*:

  Promotional materials regarding the benefits of participation in sport and physical activity had increased participants' awareness in this respect and subsequently motivated their participation. Inter- and intra-organisational competition increased motivation to take part. Creating public records of physical activity promoted commitment.

- *Those for whom positive outcomes were achieved*:

  The programme was more effectively motivational in some groups than others. Those who were previously inactive manifested the lowest impact, those who undertook occasional participation were the most likely to increase participation, and those who were the most regular participants before participating in the programme showed a slight increase in participation.

- *The contexts in which the programme was most likely to be successful*:

  The evaluation recognised that the Workplace Challenge programme was more readily embedded for those organisations in which the management team

championed employees' involvement in the programme, and in which there was 'one-to-one' direct marketing by the programme team, as compared with other organisations where such approaches were not used. Moreover, the availability of facilities (e.g., showers) at work appeared to be a potential constraint for programme participation.

In pursuing the agenda of identifying additionality, the process of systematically configuring underlying theoretical assumptions about the Workplace Challenge programme offered clear benefits. These benefits included uncovering the counterfactual scenario and attributable impacts as well as identifying the real impact of the programme. The evaluation of Workplace Challenge programme started with setting out the counterfactual scenario. Two approaches were used to assess the counterfactual scenario in order to determine whether an intervention would still be delivered in the absence of the Olympic Games: (a) by examining what is said in annual or project reports; and (b) by interviewing the programme leaders who are aware of the history of the programme. A review of programme-related documents and interviews with key stakeholders suggested that the Workplace Challenge programme would still have been delivered even if the London 2012 Games had not taken place. However, the scale of these activities, including marketing and promotion efforts, might have been negatively affected, as the programme was promoted together with some of the key sub-regionally based London 2012 events and activities. As a result, Workplace Challenge was promoted under the London 2012 banner, which might have exerted more influence on participants' decision to take part in the programme.

In terms of report-based evidence, according to the Workplace Challenge programme documents, the programme was developed as a means to increase physical activity participation at workplaces in order to increase overall physical activity participation levels and to act as a pathway to the UK Corporate Games. Even though the programme was launched after 2005 (the year when London won the Olympic bid), there is a lack of evidence to indicate that its existence was a result of the London 2012 Olympic and Paralympic Games being staged. Thus, additional information was sourced through interviewing programme stakeholders. It was specifically indicated in stakeholder interviews that the existence of the programme was not directly attributable to the 2012 Olympic and Paralympic Games. It was nevertheless recognised that, firstly, the nature of the programme delivered in 2012 was designed to be more relevant to 'London 2012'; secondly, greater prominence was given to the delivery of the programme, in terms of additional local policy and promotional support, because of London 2012.

As discussed above, the process of identifying the additionality of impacts requires us to acknowledge and operationalise the four key concepts (i.e., leakage, displacement, substitution, and multiplier effects) from the observed outcomes (English Partnership, 2008). Because of the word limit of the paper, we focus on reviewing the logic underlying the concept of leakage. The form of leakage can exist in two ways: programme beneficiaries living outside of the intended areas and benefits being generated by non-targeted groups of people (Table 2). Participants' registration information and survey results indicated that, while the majority of the programme participants resided within the intended region, there was leakage in the programme in the sense that the most

impacted group lay outside the target group: The programme had intended to encourage less active people to become more active, but a disproportionately positive response came from the already active. This process of identifying leakage was not aimed at quantifiably deducting from or increasing the results, but it nevertheless raised a point of caution when we interpreted the data, and it also proved to be useful when we were identifying those for *whom* the programme worked.

[Insert Table 2 about here]

Realist evaluation thus offered us a method for incorporating scientific[6] realist strategy into evaluation. This form of evaluation provided analytical explanation of how concepts were connected theoretically, of why there were variations between different levels of engagement, and of which organisations' own contexts enabled or disabled the effects of the designed mechanisms. Realist evaluation also provided us with the basis for a theoretically grounded understanding of what works and why (Pawson & Tilley, 2004).

## Concluding comments

In reviewing the development of evaluation research in mainstream literature, this paper has discussed specifically the four common evaluation frameworks. Particularly, it has highlighted methodological issues faced in the field of sport and discussed why realist evaluation potentially represents an advanced approach to overcoming those challenges.

Whilst it is clear that mainstream policy evaluation has moved on from simple input–output approaches to explaining policy impacts, whereby policy efficiency is assessed based on statistical association, sport policy evaluation still has some way to go.

Although we face pragmatic barriers in the field of sport, e.g. lack of research evaluation investment and limited research capacities, we shall not neglect the importance of research capabilities. We argue that it is time for the field of sport policy/programme evaluation to 'open up the black box' (Greenhalgh et al., 2009; Schmid, 1996) in order to develop causal explanations of how particular sport policy intervention can bring about the changes desired. There is still some scope for increasing the proportion of sport policies/programmes that undertake robust evaluation and for further methodological reflection in the evaluation of sport policies/programmes, and there is certainly a need for large-scale, long-term evaluations of sport programmes and initiatives (Collins, Henry, & Houlihan, 1999; Jones et al., 2017; Sanderson, 2002).

A final note of caution is that, although theory-based frameworks hold considerable promise for effectively informing policymaking, it is still a challenging task to articulate theories and subsequently measure against these causal attributions and identify causal inference factors (Chen & Henry, 2016; Granger, 1998; Sanderson, 2000). The application of realist evaluation, in practice, faces some challenges. For example, Edwards (2011) highlighted that although realist evaluation helped to deepen understanding of social policies' impacts at a micro level, it was difficult to break down the overall context into the multitude of factors associated with a specific context. When applying realist evaluation, the issues of heavy resource and labour consumption are also noted in the literature (Edwards, 2011; Gill & Turbin, 1999). Constraints concerning consumption of resources and labour could, with only a limited budget, become even more of a potential issue if both quantitative and qualitative methods were being used. While acknowledging the importance of using qualitative methods to elicit theories of

changes, Pawson and Tilly (2004) also emphasised the method-inclusive nature of the realist evaluation approach in terms of using other data collection methods (e.g., document and official records analysis).

To move forward, we would argue in favour of adopting a theory-based approach for evaluation (e.g. realist evaluation) to help with identifying the theory of change on which a sport policy/programme is based; to elucidate the causal linkages between contexts, mechanisms, and outcomes; and to allow us to begin emulating policy success and avoiding failure.

**Notes**:

[1]. Nutley, Walter, and Davies' (2007) work '*Using Evidence*', and Davies, Sandra, and Smith's (2000) book '*What works – Evidence-based policy and practice in public services*' have provided some interesting discussion of the ways in which policymakers use evidence to support their decisions and of how they judge different types of evaluations.

[2]. Potential bias may arise here as a result of programme stakeholders' control over evaluations in terms of what is looked at and how the data are interpreted, thereby compromising the scientific credibility of the evaluation. As noted by Stufflebeam and Coryn (2014), if the evaluator insists on compliance with professional standards of evaluation, the stakeholders may not welcome findings. To strike a balance between scientific creditability and fulfilling stakeholders' interests, Patton (1997) suggested that the utilisation-focused evaluation method should be carried out by highly competent and confident evaluators who have strong negotiation skills.

[3]. Readers may access more resources concerning realist evaluation and realist syntheses via the RAMESES projects (http://www.ramesesproject.org/Home_Page.php), including the quality standards and training materials developed for both approaches.

[4]. It is worth noting here that realist evaluation takes a realism philosophical approach that is different to some other forms of realism (e.g., critical realism, fallibilistic realism) (Pawson & Tilley, 1997; Pawson, 2013). Pawson argues that realist evaluation places greater emphasis on the notion of understanding the mechanics of scientific explanation and seeks to develop realism as an empirical method. In contrast, although the critical realism of Bhaskar, for example, does also agree on the importance of the classic apparatus of empirical science – such as clear conceptualisation and hypothesis-making and so on – his approach tends to focus on understanding aspects of social conditions, structures and causal powers.

[5]. Programme participants were grouped into three types according to their exercise intensity levels: Type 1 – people who were new to sport and physical activity prior to the staging of the programme, Type 2 – people who had participated in sport and physical activity but relatively less often (1–3 days a week) prior to the staging of the programme, and Type 3 – People who had regularly participated in sport and physical activity (for more than 4 days a week) prior to the staging of the Workplace Challenge programme

[6]. Here, scientific realist evaluation, as explained by Pawson and Tilley (1997), means the conducting of evaluations under the banner of *realism*.

**References**

Alkin, M. C. (Ed.) (2004). *Evaluation roots: Tracing theorists' views and influences*. Thousand Oaks, CA: Sage.

Armour, K., Sandford, R., & Duncombe, R. (2013). Positive youth development and physical activity/sport interventions: mechanisms leading to sustained impact.

*Physical Education and Sport Pedagogy, 18*(3), 256-281. doi:10.1080/17408989.2012.666791

Bailey, R., Armour, K., Kirk, D., Jess, M., Pickup, I., Sandford, R., . . . Sport Pedagogy Special Interest, G. (2009). The educational benefits claimed for physical education and school sport: an academic review. *Research Papers in Education, 24*(1), 1-27. doi:10.1080/02671520701809817

Bamberger, M., Rugh, J., & Mabry, L. (2012). *RealWorld Evaluation: Working Under Budget, Time, Data, and Political Constraints* (2nd ed.). California: Sage.

Bell, B. (2004). *An evaluation of the impacts of the Champion Coaching scheme on youth sport and coaching.* (Doctoral), Loughborough University, Loughborough.

Buisseret, T. J., Cameron, H. M., & Georghiou, L. (1995). What difference does it make? Additionality in the public support of R&D in large firms. *International Journal of Technology Management, 10*, 587-600. doi:https://doi.org/10.1504/IJTM.1995.025644

Cabinet Office. (1999). *Modernizing government, Cm.4310*. Retrieved from London:

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and Quasi-Experimental Designs for Research*. Chicago: Rand McNally and Company.

Campbell, D. T., & Stanley, J. C. (1966). *Experimental and quasi-experimental designs for research.* Skokie, IL: Rand McNally.

Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties, 2*(1), 11-20. doi:10.1017/S1745855207005029

Chelimsky, E. (2006). The purposes of evaluation in a democratic society. In I. Shaw, J. Greene, & M. Mark (Eds.), *The SAGE Handbook of Evaluation.* (pp. 33–55). London: SAGE.

Chen, H. T. (1990). *Theory-driven evaluations*. Thousand Oaks: Sage.

Chen, H. T. (2015). *Practical programe evaluation: theory-driven evaluation and the integrated evaluation perspective* (2nd ed.). Los Angeles: SAGE.

Chen, H. T., Donaldson, S. I., & Mark, M. (Eds.). (2011). *Advancing validity in outcome evaluation: Theory and practice*. San Francisco, CA: Jossey-Bass.

Chen, H. T., Grimley, D. M., Waithaka, Y., Aban, I. B., Hu, J., & Bachmann, L. H. (2008). A process evaluation of the implementation of a computer-based, health provider-delivered HIV-prevention intervention for HIV-positive men who have sex with men in the primary care setting. *AIDS Care, 20*(1), 51-60. doi:10.1080/09540120701449104

Chen, H. T., & Rossi, P. H. (1980). The multi-goal, theory-driven approach to evaluation: A model linking basic and applied social science. *Social Forces, 59*(1), 106-122. doi:10.2307/2577835

Chen, H. T., & Rossi, P. H. (1983). Evaluating with sense- The theory driven approach. *Evaluation Review, 7*(3), 283-302.

Chen, H. T., & Rossi, P. H. (1987). The theory-driven approach to validity. *Evaluation and Program Planning, 10*, 95-103.

Chen, S., Henry, I., & Ko, L.-M. (2013). Meta-evaluation, analytic logic models and the assessment of impacts of sport policies. In I. Henry & L.-M. Ko (Eds.), *Routledge handbook of sport policy* (pp. 33-47). London: Routledge.

Chen, S., & Henry, I. (2016). Evaluating the London 2012 Games' impact on sport participation in a non-hosting region: a practical application of realist evaluation. *Leisure Studies, 35*(5), 685-707. doi:10.1080/02614367.2015.1040827

Chen, S., & Henry, I. (2017). Schools' engagement with the Get Set London 2012 Olympic education programme. *European Physical Education Review, Online First*, 1356336X17721437. doi:10.1177/1356336X17721437

Coalter, F. (2007). Sports clubs, social capital and social regeneration: 'ill-defined interventions with hard to follow outcomes'? *Sport in Society, 10*(4), 537-559. doi:10.1080/17430430701388723

Coalter, F. (2010). The politics of sport-for-development: Limited focus programmes and broad gauge problems? *International Review for the Sociology of Sport, 45*(3), 295-314. doi:10.1177/1012690210366791

Coalter, F. (2011). Sports development's contribution to social policy objectives: The difficult relationship between politics and evidence. In B. Houlihan & M. Green (Eds.), *Routledge handbook of sports development* (pp. 561-578). London: Routledge.

Coalter, F. (2013). 'There is loads of relationships here': Developing a programme theory for sport-for-change programmes. *International Review for the Sociology of Sport, 48*(5), 594-612. doi:10.1177/1012690212446143

Coalter, F. (2017). Sport and social inclusion: Evidence-based policy and practice. *Social Inclusion, 5*(2), 141-149. doi:http://dx.doi.org/10.17645/si.v5i2.852

Colebatch, H. K. (1998). *Policy*. Buckingham: Open University Press.

Collins, M. I., Henry, I., & Houlihan, B. (1999). *Sport and social inclusion*. Retrieved from Loughborough:

Cook, T. D. (1995). Lessons learned in evaluation over the past 25 years. In E. Chelimsky & W. R. Shadish (Eds.), *Evaluation for the 21st century* (pp. 30-52). Thousand Oaks, CA: Sage.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings*. Chicago, IL: Rand McNally.

Cousins, J. B., & Earl, L. M. (1992). The case for participatory evaluation. *Educational Evaluation and Policy Analysis, 14*(4), 397-418.

Cronbach, L. J. (1963). Course improvement through evaluation. *Teachers College Record, 64*, 672-683.

Cronbach, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco, CA: Jossey-Bass.

Cronbach, L. J., & Associates. (1980). *Toward reform of program evaluation*. San Francisco, CA: Jossey-Bass.

Daugbjerg, S. B., Kahlmeier, S., Racioppi, F., Martin-Diener, E., Martin, B., Oja, P., & Bull, F. (2009). Promotion of physical activity in the European region: content analysis of 27 national policy documents. *J Phys Act Health, 6*(6), 805-817.

Davies, P. (1999). What is Evidence-based Education? *British Journal of Educational Studies, 47*(2), 108-121. doi:10.1111/1467-8527.00106

Davies, H. T. O., & Nutley, S. M. (2000). Healthcare: evidence to the fore. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 43-68). Bristol: Policy Press.

Davies, H. T. O., Nutley, S. M., & Smith, P. C. (2000). Learning from the past, prospects for the future. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? Evidence-based policy and practice in public services* (pp. 351-366). Bristol: Policy Press.

DCMS. (2002). *Game plan*. Retrieved from London:

De Bosscher, V., Shilbury, D., Theeboom, M., Van Hoecke, J., & De Knop, P. (2011). Effectiveness of national elite sport policies: A multidimensional approach applied to the case of Flanders. *European Sport Management Quarterly, 11*(2), 115-141. doi:10.1080/16184742.2011.559133

Dudley, D. A., Okely, A. D., Pearson, P., & Peat, J. (2010). Engaging adolescent girls from linguistically diverse and low income backgrounds in school sport: a pilot randomised controlled trial. *J Sci Med Sport, 13*(2), 217-224. doi:10.1016/j.jsams.2009.04.008

Easton, D. (1953). *The political system*. New York: Knopf.

Edwards, M. (2011). *The impact of School Sport Partnerships on primary schools: An in-depth evaluation.* (Doctoral), Durham University, Durham.

EEA. (2016). *Environment and climate policy evaluation*. Retrieved from Copenhagen:

English Partnerships. (2008). *Additionality guide - A standard approach to assessing the additional impact of interventions* Retrieved from London: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/191511/Additionality_Guide_0.pdf

Erik, A. (1995). Between knowledge and power: Utilization of social science in public policy making. *Policy Sciences, 28*(1), 79-100.

European Commission Communication. (2011). *Developing the European dimension in sport*. Retrieved from Brussels: http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2011:0012:FIN:en:PDF

Foley, P. (1992). Local economic policy and job creation: A review of evaluation studies. *Urban Studies, 29*(3-4), 557-598. doi:10.1080/00420989220080571

Furubo, J. E., Rist, R. C., & Sandahl, R. (2002). *International atlas of evaluation*. New Brunswick: Transaction Publishers.

Gabriel, K. K. P., DeBate, R. D., High, R. R., & Racine, E. F. (2011). Girls on the run: A quasi-experimental evaluation of a developmentally focused youth sport program. *Journal of Physical Activity and Health, 8*(s2), S285-S294. doi:10.1123/jpah.8.s2.s285

Gerston, L. (1997). *Public policy making: Process and principles*. Armonk, NY: M.E.Sharpe.

Gill, M., & Turbin, V. (1999). Evaluating 'realistic evaluation'. Evidence from a study of CCTV. *Crime Prevention Studies, 10*, 179-199.

Girginov, V. (2016). Has the London 2012 Olympic Inspire programme inspired a generation? A realist view. *European Physical Education Review, 22*(4), 490-505. doi:10.1177/1356336X15623169

Girginov, V., & Hills, L. (2009). The political process of constructing a sustainable London Olympics Sports development legacy. *International Journal of Sport Policy, 1*(2), 161-181.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American psychologist, 18*, 519-521.

Gortmaker, S. L., Cheung, L. W., Peterson, K. E., Chomitz, G., Cradle, J. H., Dart, H., . . . Laird, N. (1999). Impact of a school-based interdisciplinary intervention on diet and physical activity among urban primary school children: eat well and keep moving. *Arch Pediatr Adolesc Med, 153*(9), 975-983.

Granger, R. C. (1998). Establishing causality in evaluations of comprehensive community initiatives. In K. Fulbright-Anderson, A. C. Kubisch, & J. P. Connell (Eds.), *New approaches to evaluating community initiatives (volume 2): Theory, measurement and analysis*. Washington DC: Aspen Institute.

Grant Thornton, Ecorys, & Centre for Olympic Studies and Research Loughborough University. (2011a). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 1: Scope, research questions and data strategy* Retrieved from London: DCMS

Grant Thornton, Ecorys, & Centre for Olympic Studies and Research Loughborough University. (2011b). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 2: Methods*. Retrieved from London: https://www.gov.uk/government/publications/report-2-meta-evaluation-of-the-impacts-and-legacy-of-the-london-2012-olympic-games-and-paralympic-games-april-2011

Grant Thornton, Ecorys, & Centre for Olympic Studies and Research Loughborough University. (2012). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 3: Baseline and counterfactual*. Retrieved from London: DCMS

Grant Thornton, Ecorys, Loughborough University, & Oxford Economics. (2012). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 4: Interim evaluation*. Retrieved from London: DCMS

Grant Thornton, Ecorys, Centre for Olympic Studies & Research Loughborough University, & Economics, O. (2013). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 5: Post Games evaluation*. Retrieved from London: DCMS

Grant Thornton, Ecorys, Centre for Olympic Studies & Research Loughborough University, & Economics, O. (2013). *Meta-evaluation of the impacts and legacy of the London 2012 Olympic Games and Paralympic Games - Report 5: Post Games evaluation*. Retrieved from London: DCMS

Gray, A., & Jenkins, B. (1995). Policy evaluation in a time of fiscal stress: Some reflections from British experience. In R. C. Rist (Ed.), *Policy Evaluation: Linking Theory to Practice*. Aldershot: Edward Elgar.

Greene, J. (2009). Evidence as 'proof' and evidence as 'inkling'. In S. I. Donaldson, C. A. Christie, & M. M. Mark (Eds.), *What counts as credible evidence in applied research and evaluation practice?* (pp. 153-167). Thousand Oaks, CA: Sage.

Greenhalgh, T., Humphrey, C., Hughes, J., Macfarlane, F., Butler, C., & Pawson, R. (2009). How do you modernize a health service ? A realist evaluation of whole-scale transformation in London. *Milbank Quarterly, 87*(2), 391-416. doi:10.1111/j.1468-0009.2009.00562.x

Guba, E. G. (1969). The failure of educational evaluation. *Educational Technology, 9*(5), 29-38.

Guba, E. G., & Lincoln, Y. S. (1989). *Fourth generation evaluation*. Newbury Park, CA: Sage.

HallAitken. (2009). *Community sport initiative evaluation: 3rd annual report*. Retrieved from London: https://www.biglotteryfund.org.uk/-/media/Files/.../er_eval_community_sport_yr3.pdf

Hadorn, D. C., Baker, D., Hodges, J. S., & Hicks, N. (1996). Rating the quality of evidence for clinical practice guidelines. *Journal of Clinical Epidemiology, 49*(7), 749-754. doi:10.1016/0895-4356(96)00019-4.

Harris, K. (2018). Building sport for development practitioners' capacity for undertaking monitoring and evaluation – reflections on a training programme building capacity in realist evaluation. *International Journal of Sport Policy and Politics*, 1-20. doi:10.1080/19406940.2018.1442870

Henkel, M. (1991). *Government, evaluation and change*. London: Jessica Kingsley.

Henry, I. (2016). The meta-evaluation of the sports participation impact and legacy of the London 2012 Games: Methodological implications. *Journal of Global Sport Management, 1*(1-2), 19-33. doi:10.1080/24704067.2016.1177356

Hill, M. (2005). *The public policy process* (4th ed.). Pearson: Harlow.

HM Treasury. (2003). *The Green Book: Appraisal and evaluation in central government treasury guidance*. London: HM Treasury Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/220541/green_book_complete.pdf.

Hogwood, B. W., & Gunn, L. A. (1984). *Policy analysis for the real world*. Oxford: Oxford University Press.

Hood, C., Dixon, R., & Wilson, D. (2009). *'Managing by numbers': the way to make public services better*. Retrieved from http://www.publicservices.ac.uk/wp-content/uploads/policy-briefing-nov2009.pdf

Houlihan, B. (2011). Assessing the impact of sports development: Introduction - the problems of policy evaluation. In B. Houlihan & M. Green (Eds.), *Routledge handbook of sports development* (pp. 557-560). London: Routledge.

House of Sport. (2013). *Research matters: Sport Northern Ireland's research strategy 2013-2017*. Retrieved from Belfast: http://www.sportni.net/sportni/wp-content/uploads/2013/03/ResearchStrategyfinal.pdf

Hughes, K. (2013). *Sport mega-events and a legacy of increased sport participation: An Olympic promise or an Olympic dream?* (Doctoral), Leeds Metropolitan University, Leeds.

Jacob, S., Speer, S., & Furubo, J.-E. (2015). The institutionalization of evaluation matters: Updating the International Atlas of Evaluation 10 years later. *Evaluation, 21*(1), 6-31. doi:10.1177/1356389014564248

Jones, G. J., Edwards, M. B., Bocarro, J. N., Bunds, K. S., & Smith, J. W. (2017). An integrative review of sport-based youth development literature. *Sport in Society, 20*(1), 161-179. doi:10.1080/17430437.2015.1124569

Kay, T. (2009). Developing through sport: Evidencing sport impacts on young people. *Sport in Society, 12*(9), 1177-1191. doi:10.1080/17430430903137837

Keen, J., & Packwood, T. (1995). Qualitative Research: Case study evaluation. *BMJ, 311*(7002), 444-446. doi:10.1136/bmj.311.7002.444.

Leigh, A. (2009). What Evidence should Social Policymakers Use? Aus*tralian Treasury Economic Roundup*, Vol. 1, pp. 27-43, 2009. Available at SSRN: https://ssrn.com/abstract=1415462 or http://dx.doi.org/10.2139/ssrn.1415462

Lincoln, Y. S., & Guba, E. E. (1985). *Naturalistic Inquiry*. Newbury Park: CA: SAGE.

Lindsey, I., & Bacon, D. (2016). In pursuit of evidence-based policy and practice: a realist synthesis-inspired examination of youth sport and physical activity initiatives in England (2002–2010). *International Journal of Sport Policy and Politics, 8*(1), 67-90. doi:10.1080/19406940.2015.1063528

Loughborough Partnership. (2009). *School Sport Partnership Impact study*. Retrieved from Loughborough: http://www.lboro.ac.uk/departments/ssehs/research/centres-institutes/youth-sport/research/young-people-school-based/school-sport-partnership-2008.html

Mansfield, L., Anokye, N., Fox-Rushby, J., & Kay, T. (2015). The Health and Sport Engagement (HASE) intervention and evaluation project: protocol for the design, outcome, process and economic evaluation of a complex community sport intervention to increase levels of physical activity. *BMJ open, 5*(10). doi:10.1136/bmjopen-2015-009276

Marchal, B., Dedzo, M., & Kegels, G. (2010). A realist evaluation of the management of a well- performing regional hospital in Ghana. *BMC Health Services Research, 10*(1), 24. doi:10.1186/1472-6963-10-24

Marchal, B., van Belle, S., van Olmen, J., Hoerée, T., & Kegels, G. (2012). Is realist evaluation keeping its promise? A review of published empirical studies in the field of health systems research. *Evaluation, 18*(2), 192-212. doi:10.1177/1356389012442444

Mastenbroek, E., van Voorst, S., & Meuwese, A. (2016). Closing the regulatory cycle? A meta evaluation of ex-post legislative evaluations by the European Commission. *Journal of European Public Policy, 23*(9), 1329-1348. doi:10.1080/13501763.2015.1076874

McEldowney, J. (1997). Policy Evaluation and the Concepts of Deadweight and Additionality:A Commentary. *Evaluation, 3*(2), 175-188. doi:10.1177/135638909700300204

Merton, R. (1968). *Social theory and social structure*. New York: The Free Press.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of general psychology, 2*(2), 175. doi:http://dx.doi.org/10.1037/1089-2680.2.2.175

Nutley, S. M., Walter, I., & Davies, H. T. O. (2007). *Using Evidence: How research can inform public services.* Bristol: The Policy Press, University of Bristol.

Nutley, S. M., & Webb, J. (2000). Evidence and the policy process. In H. T. O. Davies, S. M. Nutley, & P. C. Smith (Eds.), *What works? - evidence-based policy and practice in public services* (pp. 13-42). Bristol: The Policy Press.

O'Brien, D. (2013). Drowning the deadweight in the rhetoric of economism: What sport policy, Free Swimming, and EMA tell us about public services after the crash. *Public Administration, 91*(1), 69-82. doi:10.1111/j.1467-9299.2011.02019.x

Österlind, M. (2016). Sport policy evaluation and governing participation in sport: governmental problematics of democracy and health. *International Journal of Sport Policy and Politics, 8*(3), 347-362. doi:10.1080/19406940.2015.1123755

Parsons, W. (1995). *Public policy: An introduction to the theory and practice of policy analysis*. Cheltenham: Edward Elgar.

Patton, M. (1984). An alternative evaluation approach for the problem-solving training program: A utilization-focused evaluation process. *Evaluation and Program Planning, 7*, 189-192.

Patton, M. (1997). *Utilization-focused evaluation-the new century text* (3rd ed.). Thousand Oaks: SAGE.

Patton, M. (2008). *Utilization-focused evaluation* (4th ed.). Thousand Oaks, CA: Sage.

Pawson, R. (2000). Middle range realism. *Archives Europeennes De Sociologie, 41*(2), 283-325.

Pawson, R. (2013). *The science of evaluation: A realist manifesto*. London: SAGE.

Pawson, R., & Tilley, N. (1994). What works in evaluation research. *British Journal of Criminology, 34*(3), 291-306.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Pawson, R., & Tilley, N. (2004). Realist evaluation. *Evaluation, 48*, 1-36.

Peirce, C. (1931). *Collected Papers* (Vol. 5). Cambridge, MA: Belknap Press.

PricewaterhouseCoopers. (2010). *Evaluation of the impact of Free Swimming_year 1 report*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/77391/FSPyear1_main_report.pdf

Ramirez, R., & Brodhead, D. (2013). *Utilization-focused evaluation: A primer for evaluators*. Retrieved from Penang, Malaysia: https://evaluationinpractice.files.wordpress.com/2013/04/ufeenglishprimer.pdf

Rist, R. C. (Ed.) (1995). *Policy Evaluation: Linking Theory to Practice*. Aldershot: Edward Elgar.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks, CA.: Sage

Rowe, N. (2005). *The value of sport*. Paper presented at the Raising the stakes. The Fourth All Island Sports Development Conference, Naas Co Kildare, Ireland.

Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation, 6*(4), 433-454. doi:10.1177/13563890022209415

Sanderson, I. (2002). Evaluation, policy learning and evidence-based policy making. *Public Administration, 80*(1), 1-22. doi:10.1111/1467-9299.00292

Schmid, G. (1996). Process evaluation: policy formation and implementation. In J. O. R. a. K. S. G. Schmid (Ed.), *International Handbook of Labour Market Policy and Evaluation*. Cheltenham: Edward Elgar.

Schoenefeld, J., & Jordan, A. (2017). Governing policy evaluation? Towards a new typology. *Evaluation, 23*(3), 274-293. doi:10.1177/1356389017715366

Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of Curriculum Evaluation* (pp. 39-83). Chicago: Rand McNally.

Scriven, M. (1969). An introduction to meta-evaluation. *Educational Products Report, 2*(5), 36-38.

Scriven, M. (1991). *Evaluation thesaurus* (4th ed.). Newbury Park: SAGE.
Scriven, M. (2008). A summative evaluation of RCT methodology and an alternative approach to causal research. *Journal of MultiDisciplinary Evaluation, 5*(9), 11-24.
Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
Smith, A., & Leech, R. (2010). 'Evidence. What evidence?': evidence-based policy making and School Sport Partnerships in North West England. *International Journal of Sport Policy and Politics, 2*(3), 327-345. doi:10.1080/19406940.2010.519341
Sport England. (2016). *Towards an active nation: Strategy 2016-2021*. London: Sport England Retrieved from https://www.sportengland.org/media/10629/sport-england-towards-an-active-nation.pdf.
Storey, D. J. (1990). Evaluation of policies and measures to create local employment. *Urban Studies, 27*(5), 669-684. doi:10.1080/00420989020080651
Stufflebeam, D. L. (1966). A depth study of the evaluation requirement. *Theory into Practice, 5*, 121-133.
Stufflebeam, D. L. (1967). The use and abuse of evaluation in Title III. *Theory into Practice, 6*, 126-133.
Stufflebeam, D. L. (1978). Metaevaluation: An overview. *Evaluation & the Health Professions, 1*(2), 146-163.
Stufflebeam, D. L., & Coryn, C. L. S. (2014). *Evaluation theory, models, and applications* (2 ed.). San Francisco, CA: Jossey-Bass, A wiley Brand.
Stufflebeam, D. L., Madaus, G. F., & Kellaghan, T. (Eds.). (2000). *Evaluation models: Viewpoints on educational and human services evaluation* (2nd ed.). Norwell, MA: Kluwer.
Suomi, K. (2004). Multiple constituency evaluation in sport policy. *European Journal for Sport and Society, 1*(2), 135-144. doi:10.1080/16138171.2004.11687754
The Lse Gv, G. (2014). Evaluation under contract: Government pressure and the production of policy research. *Public Administration, 92*(1), 224-239. doi:10.1111/padm.12055
Tilley, N. (2000). Doing realistic evaluation of criminal justice. In V. Jupp, P. Davies, & P. Francis (Eds.), *Doing criminological research* (pp. 97-113). London: Sage.
Tyler, R. W. (1942). General statement on evaluation. *The Journal of Educational Research, 35*(7), 492-501. doi:10.1080/00220671.1942.10881106
Vedung, E. (1997). *Public policy and program evaluation*. New Bruswick, NJ: Transaction Publishers.
Weed, M. (2010). How will we know if the London 2012 Olympics and Paralympics benefit health? *British Medical Journal, 340*. doi:10.1136/bmj.c2202
Weed, M. (2014). Is tourism a legitimate legacy from the Olympic and Paralympic Games? An analysis of London 2012 legacy strategy using programme theory. *Journal of Sport & Tourism, 19*(2), 101-126. doi:10.1080/14775085.2015.1053968
Weiss, C. (1972). *Evaluation*. Englewood Cliffs, NJ: Prentice Hall.
Weiss, C. (1993). Where politics and evaluation research meet. *Evaluation Practice, 14*(1), 93-106. doi:10.1177/109821409301400119

Weiss, C. (1997). How can theory-based evaluation make greater headway? *Evaluation Review, 21*(4), 501-524. doi:10.1177/0193841x9702100405

Weiss, C., & Bucuvalas, M. J. (1980). Truth tests and utility tests: Decision-makers' frames of reference for social science research. *American Sociological Review, 45*(2), 302-313. doi:10.2307/2095127

Wells, M. S., & Arthur-Banning, S. G. (2008). The logic of youth development: Constructing a logic model of youth development through sport. *Journal of Park and Recreation Administration, 26*(2). doi:http://js.sagamorepub.com/jpra/article/view/1330

Yin, R. K. (1992). The Case Study Method as a Tool for Doing Evaluation. *Current Sociology, 40*(1), 121-137. doi:10.1177/001139292040001009

Table 1: CMOs for the Workplace Challenge programme.

| Different levels of sport and physical activity (PA) participation | | | | |
|---|---|---|---|---|
| **Contexts** | **+** | **Mechanisms** | **=** | **Outcomes** |
| *People who were new to sport and PA prior to the staging of the programme*<br>e.g.<br>• The publicity surrounding the London 2012 Games made these people more aware of the benefits of sport and PA | + | e.g.<br>• Promotional materials<br>• Incentives | = | e.g.<br>• An increase in self-confidence |
| *People who participated in sport and PA relatively infrequently (1–3 days a week) prior to the staging of the programme*<br>e.g.<br>• Preferred doing sport and physical activity on their own | + | e.g.<br>• The log section<br>• Mini Challenges | = | e.g.<br>• A sense of achievement |
| *People who regularly participated in sport and PA (more than 4 days a week) prior to the staging of the programme*<br>e.g.<br>• Lack of additional free time for undertaking more sport and PA | + | e.g.<br>• Different sports competitions | = | e.g.<br>• The opportunity to try new and different types of sport and PA |
| Different categories of participating organisations | | | | |
| **Contexts** | **+** | **Mechanisms** | **=** | **Outcomes** |
| *Local authorities*<br>e.g.<br>• Worked closely with programme operators on a regular basis | + | e.g.<br>• Strong management support in their own workplaces | = | e.g.<br>• Over 50% of respondents reported that they felt ‘fitter’ and more ‘active’. |
| *Public sector*<br>e.g.<br>• Had comparatively flexible working hours, with areas for changing and shower facilities in place | + | e.g.<br>• Relatively little use of programme website | = | e.g.<br>• The programme fostered social conversation between work colleagues. |
| *Private sector*<br>e.g.<br>• Had menagerie members of staff championing employees’ involvement in the programme | + | e.g.<br>• Incentives<br>• Marketing promotion | = | e.g.<br>• Over 50% of respondents reported that their levels of sport and PA participation had increased since taking part in the programme. |

Source: Adapted from (Chen & Henry, 2016).

Table 2: Identifying the concept of leakage for Workplace Challenge.

| Concept | Definition | Programme Assumptions | Research Findings |
|---|---|---|---|
| Leakage | The extent to which the gross impact of benefits generated and intended for a particular group, region, or country incorporates beneficiaries from other groups, regions, and countries | a. As the programme was aimed at ensuring Leicestershire residents were more active, beneficiaries of the programme living outside of Leicestershire and benefiting from the programme would constitute leakage. | a. The first form of leakage proved to be negligible. Programme demographics indicated that the majority of participants resided in Leicestershire. |
| | | b. The programme was intended to target 'less active' adults in the workplace in order to increase their activity levels, but a large number of 'active' adults may in fact have benefited from the programme, and this constituted a second form of leakage. | b. There was leakage in the programme in the sense that, as identified in the interviews, the greatest impact was reported by a group of people who were already active and were therefore not the intended target (or at least not the primary intended target) of the intervention. |

Source: Adapted from (Chen & Henry, 2016).